

Mining resources

- Development of robust linguistic resources to process different types and genres of texts
 - knowledge resources about **media personalities**: recognizing and resolving references to named-entities;
 - sentiment lexicons and grammars**: detecting the polarity of opinions about media personalities
 - annotated corpora**: training different text classifiers and evaluating classification procedures

Mining resources (in progress)

- ❏ **POWER** - Political Ontology for Web Entity Retrieval
- ❏ **SentiLex** – Sentiment Lexicon
 - ❏ Human Adjectives, Nouns, Verbs and Idiomatic Expressions
- ❏ **SentiRules– Sentiment Rules**
 - ❏ Syntactic-semantic rules that determine the target of sentiment and the polarity of each predicate in context
- ❏ **SentiCorpus**
 - ❏ Sentiment annotated corpus of comments posted by the readers of the Portuguese newspaper Público to a series of 10 news articles covering the TV debates on the 2009 parliamentary election
 - ❏ Sentiment annotated corpus of tweets targeting the 5 Portuguese political leaders in the 2011 parliamentary election

SentiLex

- ▣ Describes the syntactic-semantic properties of 7,800 lemmas (80,619 inflected forms).
 - ▣ 6,319 adjectival lemmas (19,090 inflected forms)
 - ▣ 966 nominal lemmas (1,139 inflected forms)
 - ▣ 515 verbal lemmas (29,504 inflected forms)
 - ▣ 642 idiomatic expressions (30.886 inflected forms)

bem-parecido,bem-parecido.PoS=Adj;TG=HUM:N0;POL:N0=1;ANOT=MAN
desmentiram,desmentir.PoS=V;TG=HUM:N0:N1;POL:N0=0;POL:N1=-1
enterraram-se,enterrar-se.PoS=V;TG=HUM:N0;POL:N0=-1;ANOT=MAN
malvisto,malvisto.PoS=Adj;TG=HUM:N0;POL:N0=-1;ANOT=JALC
mentiras,mentira.PoS=N;TG=HUM:N0;POL:N0=-1;ANOT=MAN
mentirosos compulsivos,mentiroso
compulsivo.PoS=Adj;TG=HUM:N0;POL:N0=-1;ANOT=MAN
meteu a pata na poça,meter a pata na
poça.PoS=IDIOM;TG=HUM:N0;POL:N0=-1;ANOT=MAN
vence,vencer.PoS=V;TG=HUM:N0:N1;POL:N0=1;POL:N1=-1

REACTION

Polarity rules

☞ Lexico-syntactic patterns:

☞ [TARGET] [não|nunca] [V_{neg}|IDIOM_{neg}] > POS

e.g. Sócrates não agiu de má-fé (Sócrates did not act in bad faith)

☞ [TARGET] [V_{cop}] [ART_{ind}] [ADJ_{pos}] [ADJ_{neg}|N_{neg}] > NEG

E.g. Sócrates é um perfeito idiota (Sócrates is a perfect idiot)

☞ [o] [ADJ_{pos}] do [TARGET] > NEG

e.g. o inteligente do Sócrates (The intelligent of Sócrates)

☞ [Este|Esse] TARGET > NEG

e.g. Este Sócrates (This Sócrates)

REACTION

Polarity rules

- ☒ Lexical and “oral” clues:
 - ☒ Nicknames
 - ☒ Emoticons
 - ☒ Onomatopoeic expressions for laughter
 - ☒ Heavy punctuation marks
 - ☒ Quotation marks
 - ☒ Positive interjections

e.g. É bem provável!!! :DD RT @filipelib canção preferida de **pinócrates**? <http://bit.ly/jhpcWK> #mentirinhas

e.g. **LOOL** RT @FantasmaTioBin: O Sócrates reúne a sua equipa no balneário e grita: NÓS ATÉ O RELVAS COMEMOS :(! ""

REACTION

Sentiment Annotated Corpora

- ☒ **1,056 reader comments/4,477 sentences**
- ☒ **900 tweets**
 - ☒ Topic: Politics / Parliamentary Elections (2009,2011)
 - ☒ Annotation scheme:
 - ☒ Opinion Target & Mention
 - ☒ Opinion Polarity & Intensity
 - ☒ Opinion Literality (Literal or Ironical)
 - ☒ Opinion Holder
 - ☒ PoS (TreeTagger)

REACTION

Next steps (November 2011)

- ❏ Expand the POWER ontology to other media personalities (e.g. artists and football players).
- ❏ Enrich the POWER ontology with additional information on media personalities, crossing conventional and social media.
- ❏ Develop corpora-based approaches for automatically expanding and validating the social judgment lexicon.
- ❏ Explore new sentiment features.
- ❏ Enlarge and refine the golden collections, which will be used for training and testing text classification procedures.